of params we need to estimate (and helps lessen curse of dimensionality).

Factor Model:

- Assume $E[\vec{X}]=0$
- Assume there is a q-dimensional random vector $\vec{F}$ with $E[\vec{F}]=0$, $Var[\vec{F}]=I$, $q<p$
- $\vec{F}$ is latent/unobserved/hidden, $\vec{X}$ is manifest/observable
- Assume $\underset{p\times 1}{\vec{X}} = \underset{p\times q}{w}\,\underset{q\times 1}{\vec{F}} + \underset{p\times 1}{\vec{\varepsilon}}$, with $E[\vec{\varepsilon}]=0$, $Var(\vec{\varepsilon})=\Psi$ (Ψ = diag matrix), $Cov(\vec{\varepsilon},\vec{F})=0$

So under these assumptions

$$Cov(X_i, X_j) = Cov\left(\sum_{k=1}^{q} w_{ik}F_k + \varepsilon_i, \sum_{l=1}^{q} w_{jl}F_l + \varepsilon_j\right)$$

$$= Cov\left(\sum_{k=1}^{q} w_{ik}F_k, \sum_{l=1}^{q} w_{jl}F_l\right)$$

$$= \sum_{k=1}^{q}\sum_{l=1}^{q} w_{ik}w_{jl}\, Cov(F_k, F_l)$$

$$= \sum_{k=1}^{q} w_{ik}w_{jk} \quad \text{(since } Cov(F_k,F_l) \text{ is 0 when } k\neq l \text{ and 1 when } k=l\text{)}$$

So X_i and X_j are correlated when they load on the same factors

$Var(\vec{X}) = Var(w\vec{F} + \vec{\epsilon}) = \underset{p \times q}{w} \underset{q \times p}{w^T} + \underset{p \times p}{\psi}$

How to Estimate:

If we know ψ, then $Var(\vec{X}) - \psi = ww^T$
where ww^T is symmetric and positive definite
so $w = ud^{1/2}$ and we have
$Var(\vec{X}) - \psi = udu^T$

To estimate ψ
- regress each variable on the others and find MSE of estimates (can be initial estimate for ψ)

How to pick number of factors:

$q = rank(Var(\vec{X}) - \psi) = rank(ww^T)$
- look at eigenvalues of $\widehat{Var(\vec{X})}$ and stop when they get small (like with scree plot, and look at elbow)
- test goodness of fit, $\| \widehat{Var(\vec{X})} - (\hat{w}\hat{w}^T + \hat{\psi}) \|$ is small enough to be noise
- if we assume distribution for $\vec{F}$, we can get a dist for $\vec{X}$. Evaluate accuracy of dist using log-likelihood or something.